Manuscrit auteur, publié dans "35e congrès de l'Association Canadienne des Sciences de l'Information, Montréal : Canada (2007)"

Pierre Humbert (Prénom.Nom@loria.fr),
Equipe SITE / LORIA, Vandoeuvre-lès-Nancy, France
Institut de l'Information Scientifique et Technique INIST / CNRS,
Vandoeuvre-lès-Nancy, France

Claire François & Pascal Cuxac (Prénom.Nom@inist.fr),
Institut de l'Information Scientifique et Technique INIST / CNRS,
Vandoeuvre-lès-Nancy, France

Amos David (Prénom.Nom@loria.fr),
Equipe SITE / LORIA, Vandoeuvre-lès-Nancy, France

La visualisation des connaissances scientifiques : intégration des besoins des utilisateurs

Résumé : Cette étude porte sur l'intégration des besoins des utilisateurs dans la définition de modes de visualisation pour suivre l'évolution des connaissances scientifiques. Notre objectif consiste à optimiser le message véhiculé par les représentations graphiques et dynamiques en proposant de prendre en compte les objectifs et les attentes des utilisateurs.

Abstract : This study examines users' requirements for integrating them on visualization techniques adapted for scientific knowledge evolution. Our aim consists in optimizing the message conveyed by graphic and dynamic representations through user's goals and expectations comprehension.

hal-00274098, version 1 - 29 Oct 2008

1. Introduction

Dans le domaine de l'analyse de l'Information Scientifique et Technique, le suivi de l'évolution thématique d'un domaine scientifique est une problématique encore ouverte. La mise en évidence de thématiques émergeant des données peut être réalisée par classification en analysant soit le contenu de publications (Callon, 1993), soit les relations entre les auteurs (Small, 1993). Le problème du suivi de l'évolution de ces thématiques a été abordé dans le cadre de l'analyse des réseaux de citations des publications (Chen, 2006). Cette approche est complétée par la définition de méthodes de suivi de l'évolution thématique basées sur le contenu. En particulier, Lelu *et al.* (2006) ont défini un algorithme de classification incrémentale non supervisée dont le principe est d'intégrer au fur et à mesure les nouvelles publications d'une base de données bibliographique et de suivre les évolutions des classes résultantes qui éclatent ou fusionnent, se développent ou déclinent, apparaissent ou disparaissent.

C'est dans ce contexte que nous avons mené nos travaux de recherche. Nous nous sommes interrogés sur la manière de définir des modes de représentations graphiques qui permettent de visualiser l'évolution de ces thématiques de manière efficace pour un lecteur. Or, qu'est ce qu'une *représentation efficace* ? Cette notion n'admet-elle pas autant de réponses que d'utilisations possibles ? En effet, à partir d'un ensemble de données

partagé par des individus, chacun d'eux peut requérir un point de vue différent et fragmentaire sur ces données, selon l'objectif qu'il vise et son besoin en information.

Dans cette problématique à la croisée des domaines de la visualisation de l'information et l'ingénierie des besoins notamment, nous présenterons nos travaux en trois parties. La première partie consiste en une réflexion sur ce qui peut contribuer à l'efficacité d'une représentation graphique d'information. La seconde partie consiste en une étude des modes de visualisation dynamique de l'information, permettant de rendre compte de l'évolution des thématiques de façon lisible. Dans la troisième partie nous décrirons une mise en application dans le contexte réel de l'Institut de l'Information Scientifique et Technique (INIST) du Centre National de la Recherche Scientifique (CNRS).

2. La prise en compte du besoin, facteur d'efficacité d'une visualisation

Selon de nombreux auteurs, comme Gershon et Pages (2001) et Card *et al.* (1999), la visualisation de l'information peut se définir comme un outil de construction de sens à partir de données. La notion de *construction de sens* désigne le fait que les données laissent apparaître des phénomènes jusqu'alors invisibles lorsqu'elles sont visualisées graphiquement. Ceux-ci prennent sens pour le lecteur de manière beaucoup plus instinctive car « *les humains sont fortement habitués aux images et aux informations visuelles. Une représentation visuelle peut communiquer certains types d'information beaucoup plus rapidement et efficacement qu'une autre méthode* » (Hearst, 1999). Or, le sens que l'on peut donner à une carte, un diagramme, etc. n'aura pas la même valeur suivant ce qu'en attend le lecteur. C'est pourquoi la notion de sens nous semble étroitement lié au contexte de lecture, c'est à dire à l'objectif visé. Comme l'exprime Muchielli (2006) « *La signification (ou le sens) est à identifier à l'usage que nous faisons de l'élément en question* ».

Il n'est pas rare qu'un lecteur se sente dérouté devant une représentation, qu'il ne comprenne pas le sens de ce qu'il voit ou ne perçoive pas son utilité. Pour cela nous voyons deux raisons principales : il ne possède pas les clés de lecture de la visualisation (le mode d'emploi en quelque sorte) ou bien la représentation n'est tout simplement pas adaptée à son objectif et à son besoin d'information. Une bonne compréhension de ces deux éléments doit pouvoir améliorer l'efficacité de ce média qu'est la visualisation de l'information.

Prendre en compte le besoin d'autrui pour tenter d'y apporter une réponse de qualité la plus satisfaisante possible, est une technique courante dans de nombreux domaines. Nous citons comme exemple le cas d'une étude de marché en marketing ; la première phase de conception d'un outil informatique en génie logiciel, etc. De même l'une des fonctions de l'intelligence économique qui consiste à collecter, traiter et diffuser des informations utiles à la prise de décision dans une organisation, nécessite de prendre en compte une certaine forme de besoin, dit informationnel. En effet, un processus d'intelligence économique implique notamment deux acteurs que sont le décideur et le veilleur. Le décideur doit faire face à de nombreuses alternatives pour piloter son organisation suivant des objectifs définis. Le veilleur assiste le décideur en lui fournissant les informations dont il a besoin pour faire ses choix. Une telle interaction nécessite de la part du veilleur une attention toute particulière aux besoins informationnels du décideur.

hal-00274098, version 1 - 29 Oct 2008

hal-00274098, version 1 - 29 Oct 2008

Dans le domaine de la visualisation de l'information, nous retrouvons dans les travaux de Bertin (1967) la préoccupation de mettre en correspondance une demande d'information avec une représentation visuelle de l'information. Celui-ci développe une théorie de l'image dans le domaine de la cartographie géographique. Cette théorie avance notamment le principe d'*efficacité,* qu'il définit de la manière suivante « *Si pour obtenir une réponse correcte et complète à une question donnée, et toutes choses égales, une construction requiert un temps d'observation plus court qu'une autre construction, on dira qu'elle est plus efficace pour cette question*. ». Cette efficacité qu'il nomme aussi *prégnance*, est donc tributaire du processus de lecture, de la définition de l'image comme « *forme significative perceptible dans l'instant minimum de vision* » et des besoins informationnels que nous proposons d'identifier par une démarche d'analyse des besoins.

Dans les travaux de Tufte (2001), nous trouvons un principe nommé *principe d'excellence graphique*. Deux des cinq points qui composent ce principe le définissent comme consistant à *communiquer des idées complexes de manière claire, précise et efficiente*, mais aussi à *donner au lecteur le plus grand nombre d'idées dans le plus court intervalle de temps possible avec le moins d'encre possible dans le plus petit espace possible*. Par ce principe, la visualisation de l'information apparaît bel et bien comme un outil de communication impliquant, par conséquent, des messages. Selon Tufte, ces messages reposent sur un principe d'économie fondé sur les ressources, le temps, l'espace et l'effort du lecteur, venant ainsi corroborer le principe d'efficacité de Bertin. Cela nous conforte dans l'idée que la prise en compte des besoins informationnels des utilisateurs-lecteurs est la garantie d'une qualité et d'une efficacité des modes de visualisation.

3. Approches pour visualiser l'évolution de la science

« *En examinant des séquences chronologiques de cartes nous pouvons observer de quelle façon la connaissance scientifique évolue* » (Garfield, 1986). Depuis cette idée exprimée par E. Garfield à propos de *ISI Atlas Of Science*©, les progrès technologiques ont permis à de nombreux travaux d'être menés afin d'observer l'évolution de la science. En effet, le domaine de la scientométrie a nécessairement besoin d'un point de vue dynamique sur la science dans la mesure où son objet de recherche est en perpétuel mouvement. Quels sont les champs de recherche les plus actifs et les plus prometteurs ? Comment évoluent les réseaux d'auteurs au cours du temps ? Quels sont les points critiques de leur évolution ? Telles sont les questions pour lesquelles l'analyse de l'information peut calculer des réponses. Mais celles-ci semblent difficiles à analyser tant qu'elles ne restent accessibles qu'au travers de textes, de chiffres, de tables, etc. La visualisation peut l'y aider car, selon Card *et al.* (1999) et Gershon et Pages (2001), elle *amplifie la cognition* et « *elle permet à l'utilisateur-lecteur d'observer, de comprendre et faire sens de cette information* ».

Pour cela, Cahlik (2000) propose de visualiser les déplacements de documents parmi les thématiques obtenues par classification, à l'aide de graphes orientés. Il s'appuie sur les travaux de Callon (1993) et propose de visualiser l'évolution des thématiques au sein de diagrammes stratégiques en les comparant à différents moments. Son approche consiste à suivre l'évolution des positions des thématiques dans le diagramme. Les états des thématiques sont représentés par les sommets de graphes orientés. Chaque sommet est accompagné de sa place dans les quadrants du diagramme (en exposant) et de son

effectif (en indice). Les mouvements de documents qui s'ajoutent, transitent d'une thématique à une autre, etc. sont visibles à l'aide de flèches entrantes et sortantes.

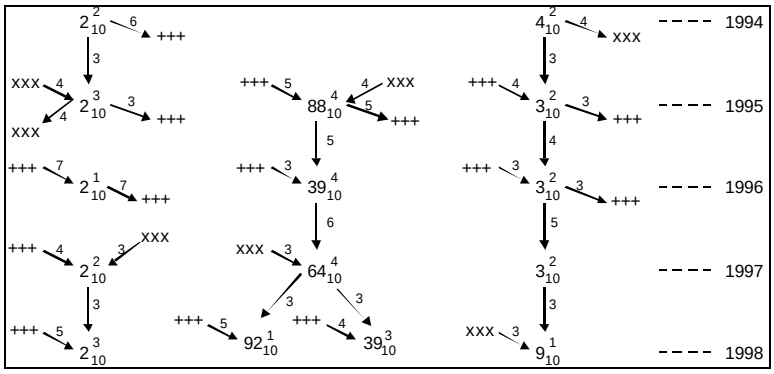

Figure 1: Évolution des classes (Cahlik, 2000)

La visualisation, telle que la reprend la figure 1, est assez rudimentaire et peu lisible : le lecteur est confronté à un ensemble relativement complexe de symboles et de chiffres. Certes ceux-ci possèdent tous une signification particulière mais l'ensemble demande un effort considérable au lecteur pour comprendre ce qu'on lui présente. Cependant, si nous la mentionnons ici c'est qu'elle permet de voir des classes qui évoluent, des documents qui transitent entre ces classes et les comportements de types éclatement notamment, sont mis en évidence (*voir* l'axe du milieu). Même si les données visualisées semblent relativement proches des données de l'algorithme incrémental sur lequel nous réfléchissons, ce mode de visualisation ne répond pas aux critères d'efficacité que nous considérons.

Graphiquement plus développée, l'approche ThemeRiver que proposent Havre *et al.* (2002), permet de visualiser des variations d'effectif des éléments constituant les thématiques. Les thématiques sont construites à partir des occurrences de termes d'un discours. L'amplitude de la courbe et la surface colorée augmente en fonction du nombre d'occurrences de ces termes. Cette représentation a l'inconvénient de ne révéler aucune structure : rien ne permet de voir des relations entre les données. Cependant, ce cas est intéressant car le lecteur perçoit bien les variations d'effectif le long de l'axe temporel, mettant périodiquement en évidence la prédominance de certaines thématiques (figure 2).

hal-00274098, version 1 - 29 Oct 2008

hal-00274098, version 1 - 29 Oct 2008

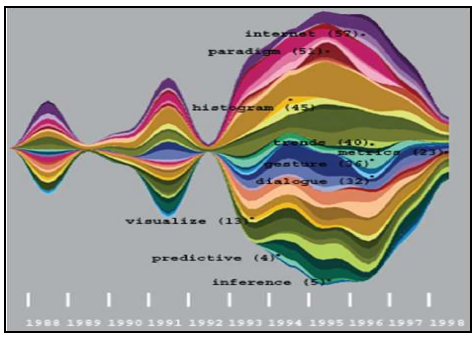

Figure 2: Visualisation de thématiques par (Havre et al. 2002)

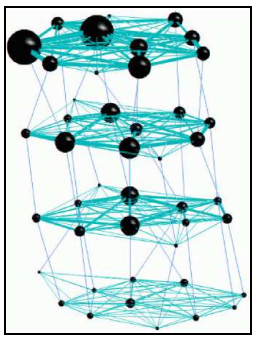

Figure 3: Visualisation de l'évolution de catégories supervisées (Erten et al., 2003)

Erten *et al.* (2003) proposent de visualiser l'évolution des thématiques grâce au système TGRIP. Celui-ci visualise l'évolution des effectifs de catégories, c'est à dire des classes supervisées de documents issus de la base bibliographique de l'ACM. Comme le montre la figure 3, la taille de chaque sommet évolue en fonction du nombre de documents que contient la catégorie représentée par ce sommet. De plus, à la différence de la représentation de ThemeRiver, cette visualisation met en évidence l'existence d'une structure entre les thématiques, même si celle-ci est peu lisible. L'évolution quant à elle est suggérée par la superposition de niveaux paraissant comme des clichés de la structure pris à différents instants. Le système a été conçu pour des catégories supervisées, aucune d'elles n'a la possibilité d'éclater ou de fusionner avec une autre, ce qui ne peut correspondre à ce que nous recherchons. Bien qu'il soit graphiquement possible de faire apparaître de tels phénomènes, cela ne nous paraît pas concevable sans altérer la lisibilité de la représentation.

Dans la continuité des travaux de Garfield sur les réseaux de citation, l'approche proposée par le système CiteSpace de Chen (2006) permet la représentation de l'évolution de ces réseaux au fil du temps. L'approche emploie deux dimensions temporelles : la date de publication et la date de citation. La date de publication de l'article détermine la place des articles, représentés par des nœuds, le long d'un axe temporel. La seconde dimension est celle qui correspond à l'année de citation : chaque nœud est caractérisé par différentes strates de couleurs qui représentent l'année ou l'article correspondant a été cité. L'ensemble, représenté sur la figure 4, permet de distinguer les *fronts de recherche* (à droite) actuels et les *bases intellectuelles* plus anciennes (à gauche).

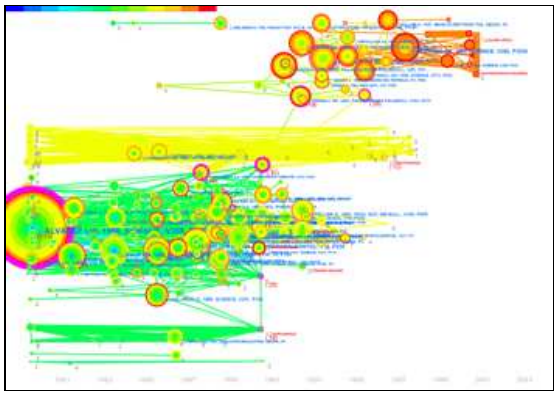

Figure 4: Visualisation de l'évolution des réseaux de citations (Chen, 2006)

hal-00274098, version 1 - 29 Oct 2008

Cette approche est fondée sur les citations et révèle une dynamique de construction de réseaux d'articles sur des thématiques proches. Le fait que cette représentation mette en évidence des structures existant entre articles, constitue un intérêt certain. Cependant, d'un point de vue graphique, là encore, l'introduction de la notion de temporalité rend la représentation et les structures elles-mêmes, difficilement lisibles. Ceci se produit lorsque le nombre de données analysées est trop important, bien que l'une des finalités d'une représentation soit justement de faciliter l'analyse de grandes quantités de données.

4. Application

L'approche que nous avons choisie pour proposer des modes de visualisation de l'information les plus efficaces possible est de prendre en compte les besoins des utilisateurs, lecteurs de cette représentation et de rechercher les éléments graphiques les plus adaptés. Nous avons appliqué cette approche dans un contexte technologique bien particulier et un cadre organisationnel spécifique.

4.1 Le contexte technologique du projet de classification incrémentale

La réflexion que nous avons mené s'est portée sur la visualisation des données obtenues en sortie d'un algorithme incrémental de classification (Lelu *et al.*, 2006), qui permet d'analyser l'évolution des publications contenues dans la base de données bibliographiques pluridisciplinaires PASCAL. Cette méthode incrémentale réalise une classification des documents par leurs descripteurs à un instant t puis prend en compte les nouveaux documents à l'instant t_{+1} et modifie les classes existantes, sans les réinitialiser à chaque instant. Il en résulte alors des classes qui se modifient, éclatent, fusionnent, apparaissent ou disparaissent. Notre problème consiste à savoir comment représenter ces données de façon lisible et efficace.

4.2 Le contexte organisationnel de l'INIST

En tant qu'organisme du CNRS, l'Institut de l'Information Scientifique et Technique, a pour fonction de collecter les publications parues dans de nombreux domaines. Avec ce matériau que sont les publications, l'INIST est appelé à accomplir plusieurs missions dont notamment donner accès à l'information scientifique et technique au public, fournir des produits de veille scientifique et technique, développer des méthodes d'analyse de l'information, etc. L'algorithme de classification incrémental, en se fondant sur les descripteurs des documents, permet de mettre en évidence des regroupements thématiques de documents qui peuvent être assimilés à des thématiques de recherche évoluant au cours du temps. Une telle approche trouve tout son intérêt dans ce contexte : perception de la dynamique de recherche scientifique pour les ingénieurs-chercheurs, suivi de l'IST pour les veilleurs et interface dynamique des bases de données pour le public.

C'est donc dans le cadre technologique du développement de cet algorithme et dans le cadre organisationnel de l'INIST que nous avons entrepris d'appliquer notre approche d'analyse des besoins afin de définir des modes efficaces de visualisation de l'évolution des classes de documents.

hal-00274098, version 1 - 29 Oct 2008

4.3 De l'identification des acteurs à la généralisation des besoins

Dans un premier temps, nous avons identifié cinq types d'acteurs ayant des besoins en information relevant du suivi de l'information scientifique et technique. Ces types d'*acteurs* sont des types d'*utilisateurs* potentiels du système exploitant l'algorithme, à savoir :

- Le responsable scientifique d'une équipe ou d'un laboratoire
- L'administrateur de la recherche
- Le chercheur
- L'étudiant ou l'historien des sciences
- Le veilleur

Dans un second temps, nous nous sommes intéressés aux problèmes auxquels ces acteurs sont confrontés, afin de repérer leurs besoins en information. Par exemple, le problème du chercheur peut être de trouver des travaux proches des siens ou d'améliorer sa visibilité, celui du responsable de recherche de définir une politique de recherche et de la situer par rapport à son environnement, celui d'un étudiant de dresser un état de l'art, etc.

Au cours du troisième temps nous avons cherché à déterminer quels éléments offerts par l'algorithme permettent de répondre à chaque problème identifié. Ce travail nous a permis de regrouper ces éléments sous 5 types de besoins informationnels :

(1) Avoir une vue globale	La plupart des utilisateurs ont besoin de positionner globalement les thématiques les unes par rapport aux autres, notamment parce qu'elles entretiennent des relations de similarité (2 thématiques proches sur la représentation ont des sujets proches). La vue globale est aussi un point d'entrée privilégié aux données comme le montre Shneiderman (2005).
(2) Observer la dynamique des thématiques	Pour un chercheur ou un responsable d'équipe qui veut améliorer sa visibilité, l'un sur ses propres travaux, l'autre sur les travaux de son équipe ; pour l'administrateur cherchant à identifier les thématiques les plus porteuses ; pour le veilleur à la recherche de signaux porteurs. Pour chacun de ces utilisateurs il est important de pouvoir observer la dynamique des apparitions, des déclins, des développements, etc. de thématiques.
(3) Retracer les étapes	Les thématiques évoluent au cours du temps, il est utile pour l'étudiant, pour l'historien ou le chercheur de pouvoir remonter la chronologie et observer les différentes étapes de leur développement.
(4) Avoir une vision transversale	Le chercheur va parfois puiser dans des thématiques différentes de celle sur laquelle il travaille, à la recherche de concepts ou de méthodes, avec des

hal-00274098, version 1 - 29 Oct 2008

	applications ou des points de vue différents.
(5) Être alerté	Le veilleur doit pouvoir percevoir les modifications significatives, être alerté des mouvements de thématiques, pour détecter les émergences et confirmer des tendances.

4.4 Définition des fonctionnalités de visualisation

Pour chacun de ces besoins, nous avons cherché à savoir quel est le mode de visualisation des informations le plus adéquat pour fournir à l'utilisateur les informations qui y répondent. Étant donné qu'une unique représentation ne peut raisonnablement couvrir tous ces besoins, nous proposons de découper la représentation en quatre approches : l'approche *globale*, *quantitative*, *qualitative*, et *par la requête* ainsi qu'en une *fonction d'alerte*.

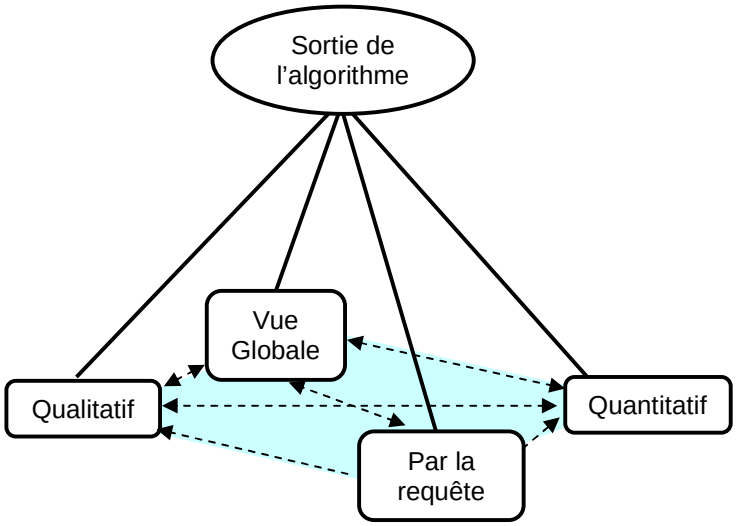

Figure 5: Quatre représentations sont nécessaires pour garantir une visualisation efficace

Les représentations que nous présentons ici sont inspirées de la littérature, nous n'avons pas encore développé le système que nous proposons ici. Sur la figure 5 qui représente ces quatre approches des résultats de l'algorithme, nous avons aussi intégré le fait que l'utilisateur puisse accéder à chacune des représentations suivant le sens matérialisé par les flèches. Nous présenterons chacune des visualisations en l'accompagnant d'une illustration tirée de l'existant et nous justifierons leur apport vis à vis des besoins identifiés.

- **La première approche est une *approche globale***. Elle permet à l'utilisateur de percevoir l'ensemble des classes et répond ainsi au besoin d'avoir une vue générale sur les données. Il s'agit d'une approche cartographique car la position des éléments de la carte est déterminée par les relations qu'ils entretiennent entre eux. La représentation inspirée par l'outil Vxinsight (Boyack *et al.*, 2002), permet cette approche globale. Nous pouvons voir les classes sur la figure 6, sous la forme d'élévations dont l'altitude est

proportionnelle au nombre d'éléments contenus. Chacune d'entre elles entretient une relation de proximité thématique avec ses plus proches voisines : plus elles possèdent des documents en commun, plus elles sont proches.

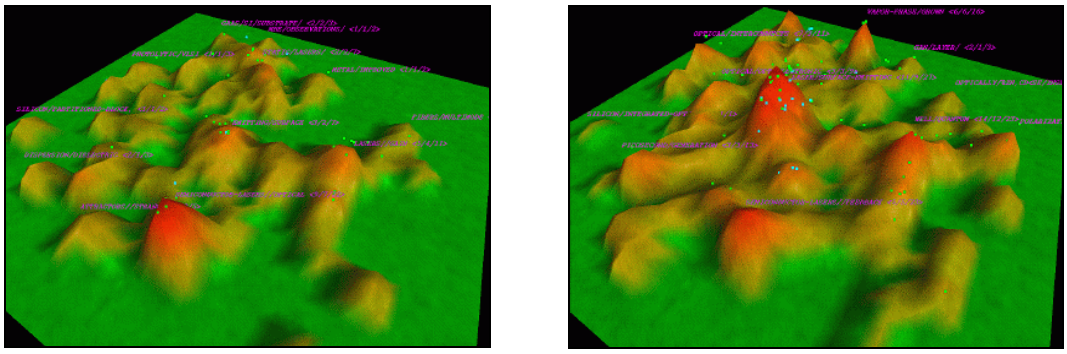

Figure 6: Une carte des thématiques puis la même carte, quelques années plus tard (Vxinsight (Boyack et al., 2002))

Cependant, cette approche ne permet pas de répondre à tous les besoins explicités. Elle ne permet pas, notamment, de déterminer l'origine des documents passant d'une classe à une autre (fusions, éclatements, etc.). Elle ne permet pas non plus de déterminer les thématiques les plus dynamiques, celles qui se développent ou déclinent dangereusement. Le rendu graphique lui-même ne permet pas d'analyser finement et avec justesse, tous les phénomènes de l'évolution calculés par l'algorithme. Par conséquent, pour mener cette analyse, il nous est nécessaire d'envisager d'autres approches.

- **La seconde approche dite *quantitative***, fondée sur l'analyse des variations d'effectifs des classes, complète la première approche. Elle permet de représenter efficacement les classes de documents qui se développent ou déclinent. Pour cela nous proposons deux représentations complémentaires. La première représentation assure le suivi de l'évolution des thématiques au cours de leur évolution à l'aide d'une courbe des effectifs en fonction du temps. Cette représentation permet de répondre à la question « *Comment se développe telle classe ?* ». Par exemple, sur la figure 7 la classe (1) a tendance à se développer tandis que la classe (2) tend à décliner.

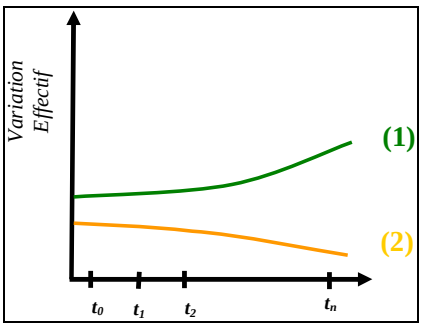

Figure 7: courbe de variation des effectifs en fonction du temps

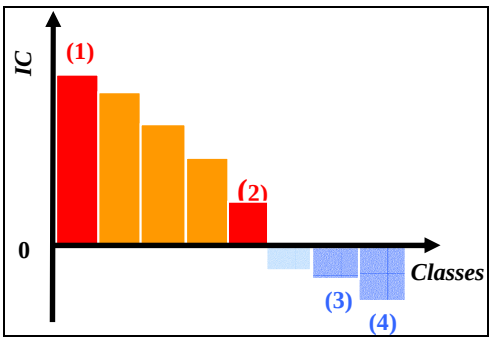

Figure 8: Représentation quantitative ordonnée fondée sur un indicateur de croissance

A cette première représentation quantitative nous pouvons en adjoindre une seconde

hal-00274098, version 1 - 29 Oct 2008

(figure 8) qui permettrait à l'utilisateur de comparer la croissance des thématiques à l'aide d'un indicateur de croissance (IC) sur une période déterminée. Cet indicateur représente l'amplitude de développement des effectifs de chaque classe entre les bornes t_{min} et t_{max} de l'intervalle de temps, par rapport à l'effectif total des documents analysés. Ainsi, une amplitude positive signifierait une croissance et une amplitude négative un déclin. L'avantage de cet indicateur est qu'il est possible d'en réaliser une représentation ordonnée comme nous le montre la figure 8. De cette manière il est envisageable de comparer les différentes dynamiques de classes, permettant à l'utilisateur de voir que la classe (1) est celle qui s'est le plus développée sur l'intervalle t_{min} à t_{max}, la classe (2) a eu un développement moyen et les classes (3) et (4) sont en déclin. Cependant, ces représentations, n'offrent pas toutes les informations dont peuvent avoir besoin les utilisateurs. Elles ne permettent pas, entre autre, de retracer l'historique d'une thématique, d'en connaître les filiations qui ont pu apparaître au cours de son évolution.

- **La troisième approche dite *qualitative*** permet de rendre compte de l'évolution des thématiques dans le temps et dans l'espace. Nous trouvons dans Tufte (2001) une représentation de la campagne de Napoléon en Russie réalisée par Minard (1869). Celle-ci retrace l'évolution des effectifs des troupes dans l'espace et dans le temps. Nous retrouvons cette représentation dans la visualisation en *spectrographe* de l'outil Autonomy (de la société du même nom), donnant la possibilité de suivre l'évolution de classes. Cette représentation permettrait, dans notre cas, de répondre aux besoins d'examiner la dynamique des classes et de retracer les étapes. Elle permet notamment de rendre compte visuellement des phénomènes d'éclatement, de fusion, d'apparition et de disparition.

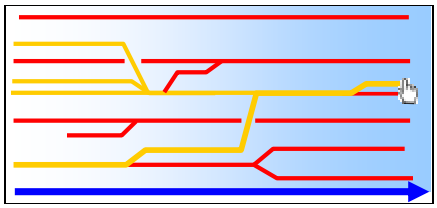

Figure 9: Historique d'une thématique (ascendances)

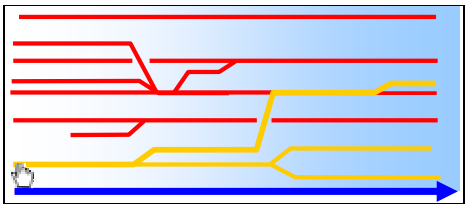

Figure 10: Affiliation d'une thématique (descendances)

Il existe cependant une limite à cette représentation : avec le nombre de classes à représenter et la multitude de comportements et donc d'entrelacements possible de celles-ci, une telle visualisation risque d'être considérablement surchargée, ce qui nuit à son efficacité. Nous proposons, pour pallier à ce problème, un mode d'interaction permettant à l'utilisateur de choisir la thématique qui l'intéresse et de visualiser l'historique, l'ascendance de celle-ci (sélection en aval, figure 9) ou son évolution, sa descendance (sélection en amont, figure 10).

- **L'approche *par la requête* est la quatrième représentation** que nous proposons. Celle-ci doit permettre à l'utilisateur d'explorer la classification à partir d'une équation de recherche. L'utilisateur ici peut donc exprimer un besoin précis portant sur le contenu des documents classifiés. Pour représenter cela nous proposons une représentation héliocentrique dont le centre serait la requête, les classes se positionnent en orbite autour de ce centre. Nous proposons aussi que la distance classes-requête soit proportionnelle à la pertinence des éléments retrouvés par le système.

hal-00274098, version 1 - 29 Oct 2008

hal-00274098, version 1 - 29 Oct 2008

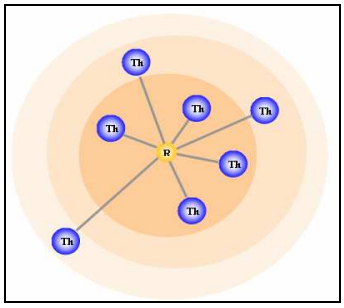

Figure 11: Visualisation héliocentrique des réponses à une requête utilisateur

Cette représentation répond au besoin d'avoir une vision transversale, puisque l'utilisateur peut visualiser toutes les thématiques plus ou moins proches de sa requête. Il serait même possible de rendre cette visualisation dynamique et ainsi permettre à l'utilisateur de suivre une évolution centrée sur la requête. Il existe en effet, de nombreux travaux portant sur le morphing de graphe qui pourraient compléter notre proposition (Loubier *et al.*, 2007).

- Enfin, dans le but de prendre en compte le besoin d'être alerté relatif à un veilleur, il est nécessaire d'introduire une dernière approche. Celle-ci prend la forme d'une fonction pouvant s'appliquer aux autres représentations, à savoir **une fonction d'alerte**. Le principe consiste à alerter le veilleur sur deux points : les variations d'effectifs et les mouvements d'effectifs. Dans le premier cas, il s'agira d'être alerté au sujet de la croissance ou du déclin des thématiques. Pour cela, l'alerte peut être fondée sur une valeur seuil qui tient compte de la progression de l'effectif de la classe par rapport à l'effectif global : une croissance est significative si l'effectif passe au delà de cette valeur (ou en deçà, pour un déclin). Dans le second cas, il s'agit d'être alerté sur le passage de documents de classe à classe. Ceci permet à l'utilisateur d'être alerté sur les fusions et les éclatements. Pour ce faire, nous préconisons l'usage de symboles correspondant aux fusions ou aux éclatements afin de favoriser ce que Bertin (1967, p.140) appelle l'*identification externe*. Ces symboles viennent se positionner sur les représentations (tracés d'évolution de l'approche qualitative, par exemple) à l'endroit où un nouveau changement est apparu.

5. Conclusion

Afin de déterminer des modes de représentation graphiques, dynamiques et efficaces, nous avons étudié en premier lieu comment nous pouvions approcher la notion d'efficacité et nous avons vu que cette notion était étroitement liée au contexte et à la notion de besoin informationnel. En second lieu nous avons parcouru des travaux issus de la littérature sur la visualisation de l'évolution de la science, pour chacun d'eux nous avons étudié leurs points faibles et leurs points forts en matière de visualisation vis-à-vis de l'objectif de notre étude. Nous avons alors formulé des propositions restées en l'état car les travaux sur l'algorithme sont en cours et aucun développement n'a été réalisé pour le moment. Nous estimons que certains points demanderaient à être approfondis : une enquête pourrait notamment venir valider les types d'utilisateurs que nous avons identifiés. Par ailleurs, dans notre étude, un type d'utilisateur est représenté par un

ensemble prédéfini de besoins. Or, celui-ci construit progressivement et individuellement des connaissances à l'aide des informations recueillies, si bien que ses besoins évoluent. Comment enrichir ce modèle en tenant compte de l'évolution des besoins individuels ? Les travaux portant sur la modélisation des comportements des utilisateurs et sur les processus d'apprentissage constituent une perspective intéressante pour répondre à ces questions (David, 1999).

Remerciements
Cette étude a été financée dans le cadre du CPER 2000-2006 pour la région Lorraine, Pôle de Recherche Scientifique et Technique – Intelligence Logicielle (PRST-IL), axe Ingénierie des Langues du Document et de l'Information Scientifique, Technique et Culturelle (ILD-ISTC).

6. Références

Bertin J. 1967. *La sémiologie graphique*. Paris : Gauthier-Villars.

Boyack, K. W., Wylie, B. N., & Davidson, G. S. 2002. Domain visualization using VxInsight for science and technology management. *Journal of the American Society for Information Science and Technology*, 53(9), 764-774.

Cahlik T. 2000. Comparison of the maps of science, *Scientometrics*, 19, n°3.

Callon M., Courtial JP, Penan H. 1993. *La scientométrie*, Paris : PUF, *Que sais-je ?*, n°2727

Card S.K., Mackinlay J.D., Shneiderman B. 1999. *Readings in information visualization using vision to think*. San Francisco : Morgan Kaufmann publisher Inc.

Chen, C. 2006. CiteSpace II: Detecting and visualizing emerging trends and transient patterns in scientific literature. *Journal of the American Society for Information Science and Technology*, 57(3), 359-377.

David, A., 1999. Modélisation de l'utilisateur et recherche coopérative d'information dans les systèmes de recherche d'informations multimedia en vue de la personnalisation des réponses. Mémoire HDR.

Erten C., Harding P.J. 2003. Kobourov S.G., Wampler K., Yee G., Exploring the Computing Literature Using Temporal Graph Visualization. Report, Department of Computer Science, University of Arizona.

Garfield E. 1986. Current Contents, 27 Octobre 1986:EIS, V.9, n°43, p.324.

Gershon N., Page W. 2001. What Storytelling can do for information visualization, In : *Communication of ACM*, Vol.44, n°8.

Havre S., Hetzler E., Whitney P., Nowell L. 2002. ThemeRiver : Visualizing Thematic Changes in Large Document Collections, *IEEE transactions on visualization and computer graphics*, Vol. 8, n°1.

hal-00274098, version 1 - 29 Oct 2008

Hearst M. 1999. User interface and visualization. In : *Modern Information Retrieval* R. Baeta-Yates, B. Ribeiro-Neto (eds), Addison-Wesley. p. 257-322

Lelu A., Cuxac P., Johansson J. 2006. Classification dynamique d'un flux documentaire : une évaluation statique préalable de l'algorithme GERMEN. In : *Journées internationales d'Analyse statistique des Données Textuelles (JADT 2006)*, Besançon, 19-21 Avril 2006, p. 617-630

Loubier E., Bahsoum W., Dousset B. 2007. Visualisation de l'évolution des informations relationnelles par morphing de graphe. In : *Journées Francophones Extraction et Gestion de Connaissances (EGC 2007)*, Namur, Belgique, 23/01/07-26/01/07, Cépaduès Editions, p. 43-54.

Minard, C. J. 1869. Carte figurative des pertes successives en hommes de l'armée française dans la campagne de Russie, 1812-1813. lith. (624 x 245), 20 novembre 1869. ENPC: Fol 10975, 10974/C612.

Mucchielli A. 2006. *Le dialogue avec la technologie,* Paris : Armand Colin.

Shneiderman B., Plaisant C. 2005. *Designing the user interface, Strategies for effective human-computer interaction,* 4e édition. Addison Wesley.

Small H. 1999, Visualizing Science by Citation Mapping, *Journal of the American Society for Information Science,* n°50(9), pp. 799-813.

Tufte R. E. 2001. *The visual display of quantitative information*, Cheshire : Graphic press.

hal-00274098, version 1 - 29 Oct 2008